

Fire Detection in Color Images using Markov Random Fields

David Van Hamme¹², Peter Veelaert², Wilfried Philips¹, and Kristof Teelen¹²

¹ Ghent University/IBBT (IPI)

² University College Ghent (Vision Systems)

Abstract. Automatic video-based fire detection can greatly reduce fire alert delay in large industrial and commercial sites, at a minimal cost, by using the existing CCTV camera network. Most traditional computer vision methods for fire detection model the temporal dynamics of the flames, in conjunction with simple color filtering. An important drawback of these methods is that their performance degrades at lower frame-rates, and they cannot be applied to still images, limiting their applicability. Also, real-time operation often requires significant computational resources, which may be unfeasible for large camera networks. This paper presents a novel method for fire detection in static images, based on a Markov Random Field but with a novel potential function. The method detects 99.6% of fires in a large collection of test images, while generating less false positives than a state-of-the-art reference method. Additionally, parameters are easily trained on a 12-image training set with minimal user input.

1 Introduction

Fire detection is an important component of industrial and commercial site surveillance systems with regard to personnel and material safety. Nearly all of the currently employed systems rely on dedicated sensors and manually activated fire alarms. To detect fire as early as possible, a combination of different sensor types is often made, linked by sensor fusion methods to improve reliability. Examples of such techniques include Bao *et al.* [1], who use temperature and photo-electric smoke sensors, and Li *et al.* [2], whose techniques rely on multi-spectral cameras. These systems however, are impractical or too expensive for covering large sites, especially outdoors, due to the required sensor density. A cheap and effective alternative is the use of computer vision-based techniques in conjunction with digital cameras or CCTV networks. The main advantages are the large coverage area offered by a single camera, and the possibility of integration with existing surveillance camera systems.

The state-of-the art fire detection methods in computer vision typically consist of two main parts, modelling the most characteristic aspects of fire in video. The first aspect is spectral information. All methods employ a color filter of some sort, usually based on a fixed set of rules. The second aspect concerns the temporal dynamics of flames, often combined with spatial characteristics. The

spatio-temporal modelling of fire in video was first described by Healey *et al.* [3] in 1993, with more recent contributions by Liu *et al.* [4] and Töreyn *et al.* [5]. Other examples of temporal properties of fire used in fire detection include standard background subtraction [6], flame growth and propagation [7, 8], intensity and boundary flicker [9], area, roundness and circumference deviation [10], edge dynamics [11] and temporal contour analysis [12].

An important limitation of all these methods is that their performance degrades at lower framerates, as accurate modelling of flame dynamics requires a high temporal resolution, and they cannot be applied to still images. Additionally, the high video data rates combined with the requirement of real-time operation mean that significant computational resources are necessary to monitor a single video stream. This is an important concern for large camera networks. These two drawbacks also inhibit the use of the methods for low-power, wireless camera systems, where frame rates are low to save transmission time and thus save battery power, or where processing is integrated in the camera itself. Some efforts have been made to produce fire detection systems for still images, notably by Noda *et al.* [13], who employed color histogram models for tunnel security monitoring. For use in a more general setting, the static color filters used in the dynamic methods can be used and improved upon, but they still yield a high false alarm rate [14]. Also, as the filters rely on a predefined set of rules, they require time-consuming parameter tuning.

In this paper, a novel method is presented for fire detection on static images. Rather than using a set of rules in color space, the image data is treated as a Markov Random Field (MRF). MRF theory is a powerful tool for modeling contextual dependencies, and has successfully been applied to a variety of texture classification problems [15, 16]. The MRF we propose employs a custom potential function shaped by training data. A classifier evaluates the energy function of the MRF per image block to detect blocks on the border of flames. The method is shown to yield near perfect detection rates on a variety of fires, while generating less false alarms than a state-of-the-art fixed-threshold color filter.

2 The MRF model

Markov Random Fields theory is a branch of probability theory developed for modeling contextual dependencies in physical phenomena. In computer vision, it is primarily used for labeling problems, to establish probabilistic distributions of interacting labels. A thorough description of the application of MRFs to vision problems can be found in the book “MRF Modeling in Computer Vision” by S. Z. Li [17]. The basic principles, terminology and notation are described below.

Let $\mathbf{S} = \{i | i = 1 \dots m\}$ be an index set corresponding to a set of sites in a Euclidian space (e.g. a regular two-dimensional lattice), in which each site is uniquely defined by its index i , and let \mathbf{L} be a discrete or continuous set of labels. Let $\mathbf{F} = \{F_1, \dots, F_m\}$ be a family of random variables defined on \mathbf{S} , in which each random variable F_i takes a value from a label set \mathbf{L} . The label of the random variable F_i will be denoted f_i . Assuming a discrete label set, the

probability that F_i takes on a certain label f_i is given by $P(f_i)$. The family \mathbf{F} is called a random field on \mathbf{S} . The joint probability of the random field taking a particular combination of values is denoted $P(\mathbf{f})$.

A Markov Random Field is defined as a random field in which the probability $P(f_i)$ is only dependent on f_i and some of its neighbors. Therefore, a neighborhood system \mathbf{N} is defined as

$$\mathbf{N} = \{\mathbf{N}_i | i \in \mathbf{S}\} \quad (1)$$

where \mathbf{N}_i is the index set of sites neighboring i . The neighboring relationship has the following properties:

1. a site is not a neighbor of itself: $i \notin \mathbf{N}_i$,
2. the neighboring relationship is mutual: $i \in \mathbf{N}_{i'} \iff i' \in \mathbf{N}_i$.

For a regular lattice \mathbf{S} , the neighboring set of i is usually defined as the set of sites within a radius of i . Note that sites at or near the boundary of the lattice have fewer neighbors. The Markovianity constraint is then expressed by

$$P(f_i | \mathbf{f} - \{f_i\}) = P(f_i | f_{\mathbf{N}_i}) \quad (2)$$

where $\mathbf{f} - \{f_i\}$ denotes all values of the random field except for f_i itself, and $f_{\mathbf{N}_i} = \{f_{i'} | i' \in \mathbf{N}_i\}$ stands for the labels at the sites neighbouring i .

Let us construct a graph on \mathbf{S} in which the edges represent the neighboring relationships. Now consider the cliques in this graph. A clique is a subset of vertices so that every two vertices are connected by an edge. In other words, the cliques represent sites which are all neighbors to each other. Thus, a clique consists of either a single site, or a pair of neighboring sites, or a triple, and so on. The collection of single-site and pair-site cliques will be denoted by \mathbf{C}_1 and \mathbf{C}_2 respectively, where

$$\mathbf{C}_1 = \{\{i\} | i \in \mathbf{S}\} \quad (3)$$

$$\mathbf{C}_2 = \{\{i, i'\} | i' \in \mathbf{N}_i, i \in \mathbf{S}\}. \quad (4)$$

The energy function $U(\mathbf{f})$ is a measure of the likeliness of the occurrence of \mathbf{f} for a given model. For single-site and pair-site cliques, it is defined as

$$U(\mathbf{f}) = \sum_{i \in \mathbf{S}} V_1(f_i) + \sum_{i \in \mathbf{S}} \sum_{i' \in \mathbf{N}_i} V_2(f_i, f_{i'}) \quad (5)$$

where V_1 and V_2 denote potential functions for single-site and pair-site cliques. Lower energy of the joint distribution represents a better fit of the model to the data.

When applied to digital images, the sites correspond to pixel locations, and the neighborhood system is usually either 4-connectedness or 8-connectedness. For a 4-connected system, the four types of pair-site clique that any non-edge pixel belongs to are shown in figure 1.

A type of MRF of particular interest to labeling problems in computer vision is the Multi-Level Logistic (MLL) model. In an MLL, the potential functions are defined as

$$V_1(f_i) = \alpha_{f_i} \quad (6)$$

where α_{f_i} is the potential associated with the label f_i , and

$$V_2(f_i, f_{i'}) = \begin{cases} \beta & |f_i = f_{i'} \\ -\beta & |f_i \neq f_{i'} \end{cases} \quad (7)$$

where β is the potential for pair-site cliques. For the 4-connected neighborhood system, each non-edge pixel belongs to four different pair-site cliques, as shown in figure 1. The reason why this model is often used in computer vision, is that for $\beta < 0$, the MLL model acts as a smoothness prior. The potential function then favors smooth distributions with blob-like regions of uniform labels, which is a desirable property in labeling algorithms.

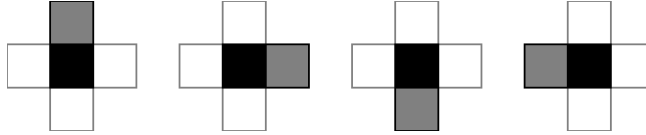


Fig. 1. The four types of pair-site cliques any non-edge pixel belongs to in a 4-connected neighborhood. Black is the pixel concerned, gray is the additional pixel that makes up the clique, outlined in gray is the neighborhood.

For our fire detection application, we want to use the MRF to model the typical red to yellow color texture found in flames. The label set \mathbf{L} therefore consists of discrete color labels, obtained by binning the hue channel in HSV color space. The hue is divided into n evenly spaced bins, with $n \geq 6$, as separation of yellow and red are essential. typically, $n = 8$. Now we must choose the potential functions V_1 and V_2 so that the joint energy $U(\mathbf{f})$ is low for fire areas and high for background. There are two properties of fire we would like to model in the MRF. The first property is the typical color range of the fire pixels. This can be expressed in the potential function for single-site cliques, V_1 , by choosing the values of α_{f_i} low for typical fire colors and high for the others. The second property we want to model is spatial hue variation, reflecting the color gradients typically present in flames. This property will help set apart actual fires from uniformly fire-colored objects, e.g. fire trucks, billboards or fire-colored clothing items. This *unsmoothness* prior can be implemented in the potential function for pairwise cliques, V_2 , by specifying a positive value for the constant β .

A straightforward way to construct a classifier from this model is to evaluate the joint energy $U(\mathbf{f})$ per 4×4 pixel block of the image, and setting a threshold

on this energy below which the block is classified as belonging to fire. Our experiments have shown that this technique works and can produce adequate detection rates. However, the false positive rates do not display a significant improvement over state-of-the-art color-based methods [14]. This can be attributed to the simple color space binning. Since V_2 favors any kind of label variation, pair-site cliques consisting of different non-fire colors will also generate low energy. This limits the usefulness of the potential function V_2 and shifts the importance towards V_1 , thereby diminishing the advantage the MLL should theoretically offer. In practice, this means after optimizing the parameters, β becomes insignificant compared to the values α_{f_i} . A solution to this problem is given in the next section.

3 A custom potential function

One way to resolve the issue described above, would be to adapt the color segmentation to obtain a more intelligent labeling. However, this means dealing with an issue we are trying to avoid as much as possible: setting hard thresholds in color space. As an alternative way to improve the false positive rate, we propose a novel potential function. Rather than specifying a constant value for β , it will now depend on the relative occurrence of the particular clique in a foreground (fire) and background model. Note that this means a departure from the MLL theory. For every possible pair of color labels, a potential value is now calculated beforehand, based on training data. Let $C_f(f_i, f_j)$ denote the number of times a pairwise clique consisting of the labels f_i and f_j occurred over all fire areas in the training data, and likewise $C_b(f_i, f_j)$ the number of times it occurred over all background areas. We will then estimate the occurrence probabilities of the clique in foreground and background as

$$P_f(f_i, f_j) = \frac{1 + C_f(f_i, f_j)}{\sum \sum_{f_i, f_j \in L} C_f(f_i, f_j) + |L|^2} \quad (8)$$

$$P_b(f_i, f_j) = \frac{1 + C_b(f_i, f_j)}{\sum \sum_{f_i, f_j \in L} C_b(f_i, f_j) + |L|^2} \quad (9)$$

Note that we added 1 to the occurrence counts of each clique to avoid probabilities of zero, as is common practice (e.g. for training a Bayes classifier). This gives rise to the term $|L|^2$ in the denominator. The potential function V_2 we propose is then given by

$$V_2(f_i, f_{i'}) = \begin{cases} \frac{P_b(f_i, f_j)}{P_b(f_i, f_j) + P_f(f_i, f_j)} & |f_i = f_j \\ -\frac{P_f(f_i, f_j)}{P_b(f_i, f_j) + P_f(f_i, f_j)} & |f_i \neq f_j \end{cases} \quad (10)$$

The value in the first case is the probability that, if this particular clique occurs, it is caused by the background model. Likewise the value in the second case is the probability that it is caused by the fire model. While this potential function is obviously heuristic, it implements the functionality we require:



Fig. 2. Source frame, detected blocks with high energy threshold, and corresponding component with lower energy threshold.

- cliques of uniform color are penalized, but more so for unlikely fire colors,
- cliques of different color are encouraged, but more so for typical fire combinations.

Experiments show that with the new potential function V_2 , the areas near the edges of flames generate very low energy, while the entire background results in much higher energy values. The interior part of the flame falls in between, on average generating more energy than the flame edge but less than the background. This is exploited in a two-stage classifier. The energy function is calculated on 4x4 pixel blocks and first thresholded on an energy level T_1 allowing the entire fire areas to pass the criterion, as well as some spurious detections in the non-fire areas. In the second stage, the calculated energy is thresholded on a level $T_2 < T_1$, allowing only series of blocks near the edges of the flames. Only the first-stage connected components which contain blocks from the second stage are retained, resulting in much fewer false detections. An example can be seen in figure 2.

4 Performance

| Method | Detection rate | False alarm rate |
|-----------------------------------|----------------|------------------|
| Celik 2008 | 99.95% | 50.93% |
| Proposed (first stage only) | 99.98% | 42.82% |
| Proposed (first and second stage) | 99.57% | 21.80% |

Table 1. Performance statistics of Celik *et al.* compared to the proposed method.

The color occurrence probability distributions P_f and P_b were trained on a set of 6 ground truth images of fire, and 6 additional background images featuring a variety of settings. The fire images are video frames depicting four different fires, captured by different types of cameras and from different viewing

angles. The images also exhibit a wide variety of camera settings, from underexposure to oversaturation and varying degrees of focal sharpness. The additional background images were included for training balance, as the video frames were all captured in an industrial environment and therefore featured similar backgrounds. This training is intended to be universal, so no retraining is required for use in different circumstances. However, results may improve further for very specific scenarios when the method is trained on the according scenario-specific imagery.

The performance of the fire detection system was evaluated on over 49,000 video frames and compared to the fire detection method proposed by Celik *et al.* [14], which defines a set of rules in Cr-Cb color space based on three polynomial curves. The method was implemented as described in the paper, taking care to use the same 8-bit range for the chroma planes. The results were also aggregated into 4x4 blocks using a majority voting rule, to make comparison with our method as fair as possible. We consider this method to be the state-of-the-art single-pixel fire color filter against which to judge the benefits of our contextual modeling.

The first part of the test set consists of 30,000 frames depicting fire, to obtain the detection rate. These video frames show a number of controlled fires in an outdoor firemen training complex built to resemble an industrial site. The fires include a burning petroleum tank, a ruptured gas pipe, a round tank engulfed in flames and a fire in a maintenance trench. The fires were monitored by six cameras of different types, placed on different elevation levels and angles. The fire is considered detected as soon as at least one of its pixel blocks is detected as a fire block. In the interest of fairness, we should note that the training images for our method were captured on the same site, albeit at a different time with different sunlight levels.

The second part of the test set contains over 18,000 video frames captured from a moving vehicle in an urban environment. This set is representative of the occurrence of fire-colored objects to be expected in the busiest environments, e.g. red and yellow clothing, vehicles or advertising. An image is counted as a false positive when one or more blocks in the image are classified as fire.

The results obtained on this data set are shown in Table 1. The reference method by Celik *et al.* scored a detection rate of 99.95% on the fire frames, while generating over 50% false positives. This illustrates the high occurrence of fire-colored objects in the second dataset: over half of the frames contain at least one fire-colored 4x4 block. In comparison, the detection rate of our proposed method after just the first stage was 99.98%, with a false detection rate of 42.82%. This shows that even after just the least discriminative of the two stages, there is an improvement over the reference method. After both stages of the method, the detection rate drops only slightly to 99.57%, while false positives are much reduced to 21.80%. These statistics prove the adequacy of the system as standalone fire detector. The cases in which the fire was not detected are mostly transition phases, either just after the fire was started or when it was nearly extinguished. One can reasonably assume that any spreading fire will be detected. The false

negatives can thus be considered rare and temporary manifestations of fire in which the spectral texture is coincidentally and atypically low.

5 Conclusion

We have designed an MRF-based visual fire detection system which is easy to train, and requires optimization of just one critical parameter (the lower classifier energy threshold) rather than setting multiple fixed color rules. Furthermore, after training on basic, generic ground-truth data the method is proven to yield very good detection rates in a variety of circumstances, while at the same time significantly reducing false positives over standard color-based methods. Moreover, it does not rely on any temporal information, and can therefore be applied to still images and low framerate cameras without performance degradation. On the other hand, if normal video frame rates and sufficient computing power are available, the method could be improved further by implementing temporal hysteresis, whereby multiple subsequent alerts are required before the alarm is set off. The model uses insignificant amounts of memory (typically 256 bytes) and the block-based processing suits parallel implementation, making the method ideal for implementation on dedicated hardware (e.g. FPGAs) to speed up computation.

References

1. Bao, H., Li, Y., Zeng, X.Y., Zhang, J.: A fire detection system based on intelligent data fusion technology. In: Machine Learning and Cybernetics, International Conference on. Volume 2. (2003) 1096–1101
2. Li, Y., Vodacek, A., Kremens, R.L., Ononye, A., Tang, C.: A hybrid contextual approach to wildland fire detection using multispectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on* **43** (2005) 2115–2126
3. Healey, G., Slater, D., Lin, T., Drda, B., Goedeke, A.: A system for real-time fire detection. In: CVPR. (1993) 605–606
4. Liu, C.B., Ahuja, N.: Vision based fire detection. In: ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4, Washington, DC, USA, IEEE Computer Society (2004) 134–137
5. Töreyn, B.U., Dedeoglu, Y., Gündükbay, U., Çetin, A.E.: Computer vision based method for real-time fire and flame detection. *Pattern Recogn. Lett.* **27** (2006) 49–58
6. Celik, T., Demirel, H., Ozkaramanli, H., Uyguroglu, M.: Fire detection using statistical color model in video sequences. *J. Vis. Comun. Image Represent.* **18** (2007) 176–185
7. Huang, P.H., Su, J.Y., Lu, Z.M., Pan, J.S.: A fire-alarming method based on video processing. *Intelligent Information Hiding and Multimedia Signal Processing, International Conference on* (2006) 359–364
8. Ko, B., Hwang, H.J., Lee, I.G., Nam, J.Y.: Fire surveillance system using an omnidirectional camera for remote monitoring. In: CITWORKSHOPS '08: Proceedings of the 2008 IEEE 8th International Conference on Computer and Information Technology Workshops, Washington, DC, USA, IEEE Computer Society (2008) 427–432



Fig. 3. Examples frames and their detector output.

9. Töreyn, B.U., Çetin, A.E.: Online detection of fire in video. In: CVPR. (2007)
10. Zhang, D., Shizhong, H., Zhao, J., Zhang, Z., Chengzhang, Q., Youwang, K., Xiang, C.: Image based forest fire detection using dynamic characteristics with artificial neural networks. In: Artificial Intelligence, International Joint Conference on. (2009) 290–293
11. Kandil, M., Salama, M.: A new hybrid algorithm for fire vision recognition. In: EUROCON 2009, IEEE. (2009) 1460–1466
12. Zhang, Z., Zhao, J., Zhang, D., Qu, C., Ke, Y., Cai, B.: Contour based forest fire detection using fft and wavelet. In: CSSE '08: Proceedings of the 2008 International Conference on Computer Science and Software Engineering, Washington, DC, USA, IEEE Computer Society (2008) 760–763
13. Noda, S., Ueda, K.: Fire detection in tunnels using an image processing method. In: Vehicle Navigation and Information Systems Conference, Proceedings. (1994) 57–62
14. Celik, T., Kai-Kuang, M.: Computer vision based fire detection in color images. In: Soft Computing in Industrial Applications, IEEE Conference on. (2008) 258–263
15. Xiang, Y., Zhou, X., Chua, T., Ngo, C.: A revisit of generative model for automatic image annotation using markov random fields. In: CVPR. (2009) 1153–1160
16. Chan, A., Vasconcelos, N.: Layered dynamic textures. Pattern Analysis and Machine Intelligence, IEEE Transactions on **31** (2009) 1862–1879
17. Li, S.Z.: Markov Random Field Modeling in Computer Vision. Springer-Verlag (1995)